Apache Kylin在百度地图的实践

百度地图开放平台业务部 - 王冬



大数据多维分析平台建设背景

大数据查询框架选型及调研

大数据OLAP多 维分析平台建设

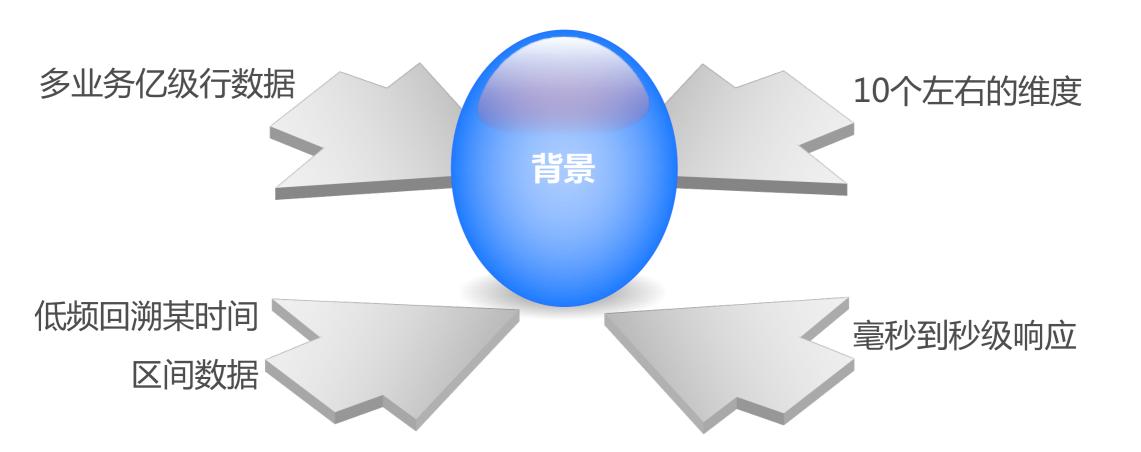
我们的深度开发

我们的项目实践

讨论



大数据多维分析平台建设背景





大数据OLAP多维分析平台建设



大数据查询框架选型及调研 - 相关生态

Metadata, SQL Parser 离线计算及查询 -实时 全文 机器 **Job Engine. Optimizator, Query Engine Storage Engine** 计算 检索 学习 交互式查询/用户推送/广告/Adhoc 实时 挖掘 检索 数据应用 ETL 查询 推荐 查询等 应用 数据仓库 大数据OLAP多维分析平台 数据服务 (分钟、小时级响应) (毫秒、秒级查询响应) Storm, Spark Elastic Spark Mllib, Search, **Apache** Hive(+Tez framework) Streaming, Spark SQL Mahout Lucene **Kylin** Flink **Impala** 数据计算 Hadoop **HBase DataFrame** 数据缓存 Tachyon(in memory file system) Hadoop Distributed File System(HDFS) 数据存储 Scribe **Flume** 数据引入 **HDFS JSON** Text **Parquent** Other 数据源 **Data Source Data Source Data Source Data Source Data Source**

大数据查询框架选型及调研 - 技术选型

MySQL: 维度少+数据量级小+查询响应慢 1.0 松地震 2.0 大数据计算查询 3.0 为这间查询

大数据业界离线计算查询框架分类

基于Runtime Framework的架构 (如Hive, Pig)(分钟、小时级)

基于内存计算的MPP架构

(如Impala, Presto, Spark SQL, Palo) (秒级、分钟级)

基于预计算和索引的架构

(如Apache Kylin, Pinot)(毫秒、秒级)



大数据OLAP多维分析平台建设



大数据OLAP多维分析平台建设 - 系统位置

业务线用户

(需求方,平台使用方)



数据仓库 (udw, biglog)

OLAP多维分析

(Apache Kylin , Impala, Spark SQL, Presto)

前端可视化

(Smart DI, Saiku , Tableau, Qlikview)

第一步

业务方提出需求,多方 共同审批,并在项目开 发完毕后进行数据验证。 上线后,平台的主要使 用者。

第二步

日志入库udw,仓库端利用biglog平台,进行项目最细粒度维度下,指标的逻辑开发

第三步

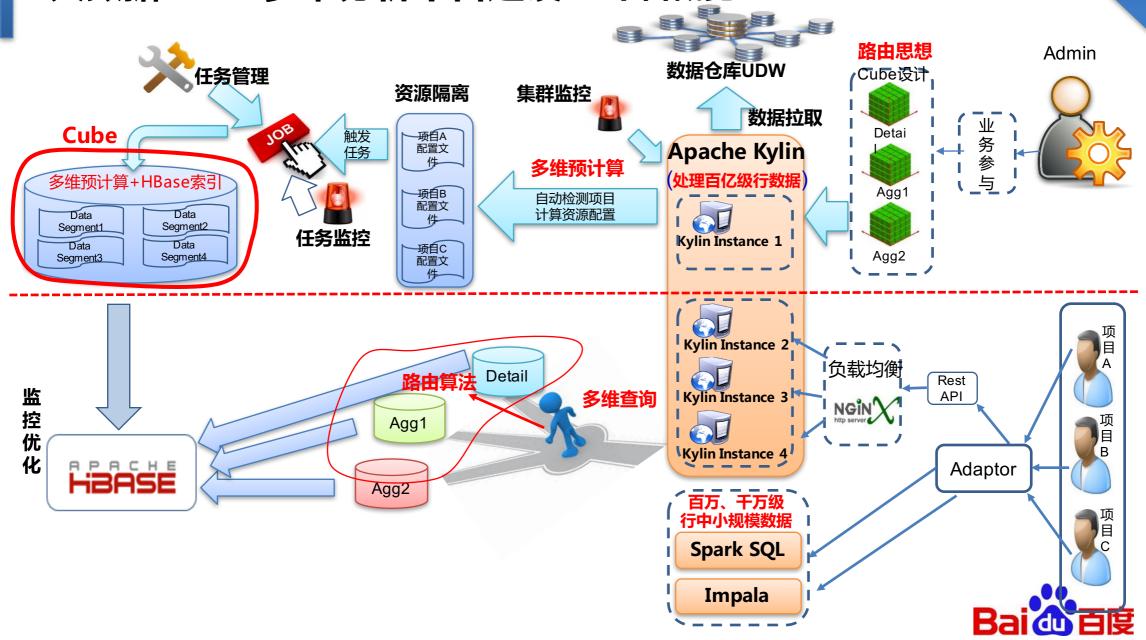
亿级行数据,基于内存或基于预计算及索引, 毫秒级、秒级交互式多维分析查询。

第四步

自动化程度较高的前端 配置或者拖拽生成的多 维分析页面



大数据OLAP多维分析平台建设 - 平台概览



大数据OLAP多维分析平台建设 – 平台服务

旧系统 新系统

系统

基于MySQL等引擎

基于Memory-Based引擎

基于Pre-calculated引擎

支持数据量级

x <百万行级

百万行级<x<亿行级

千万行级 < x < 百亿行级

支持维度数量

小于5个

任意

<=15个(建议)

支持指标数量

小于5个

任意

任意

查询响应

毫秒到秒

秒到分钟

毫秒到秒

系统特点

RDBMS

内存

空间换时间



我们的深度开发



我们的深度开发 – 开发工作

(((Q)))

数据拉取

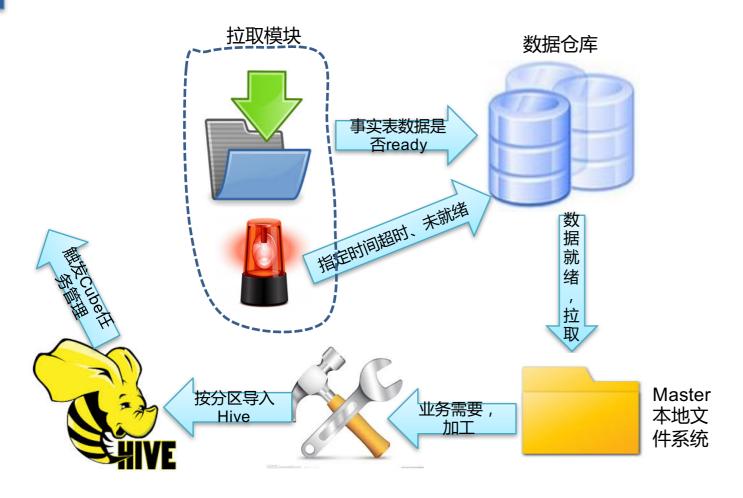
平台部署 监控优化 任务调度 # 1 Hadoop 存储原则 Hadoop优化 数据拉取依赖监控 Hive 计算任务 Hive优化 Cube任务监控 HBase优化 **HBase** 更新任务 Hadoop,HBase监控 Zookeeper优化 Kylin 合并任务 Apache Kylin监控 01 02 03 04 05 异构数据源 依赖检测 项目粒度 资源隔离 超时预警 落地加工

资源隔离

Bai d 首度

原生态暂不支持资源隔离

我们的深度开发 – 数据拉取



数据源: HDFS, MySQL等

依赖检测: 数据就绪依赖检测

超时预警: 数据在指定时间内未就绪,

短信或邮件报警

落地加工: 根据业务需要,加工落地

的数据,再导入Hive并计算。



我们的深度开发 - 任务调度

数据存储原则:

当月数据保持天为单位的分区, 历史数据保持月为单位的分区

计算(build)

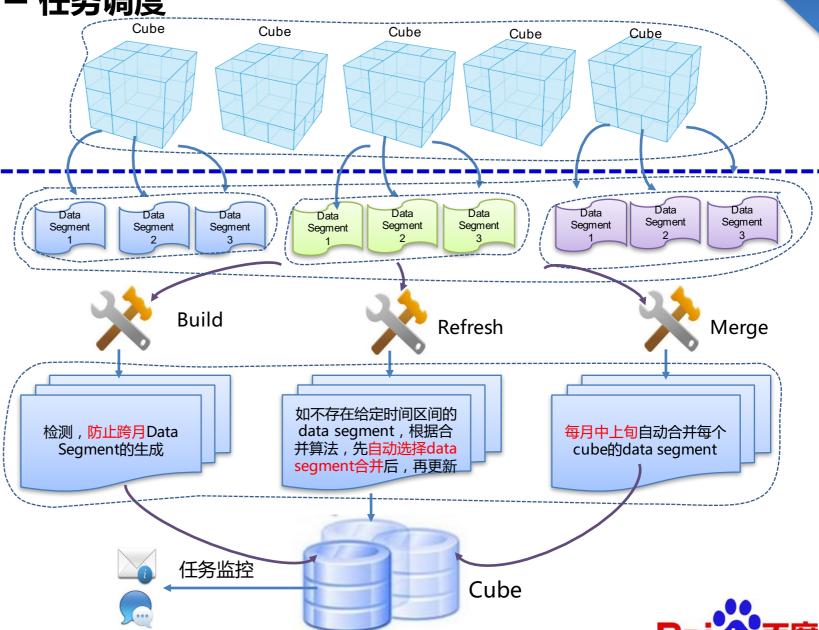
自动划分长时间线段成自然月。 源数据延迟产出可能造成数据分 区跨月的问题

更新(refresh)

当月数据回溯及历史数据优化回 溯策略

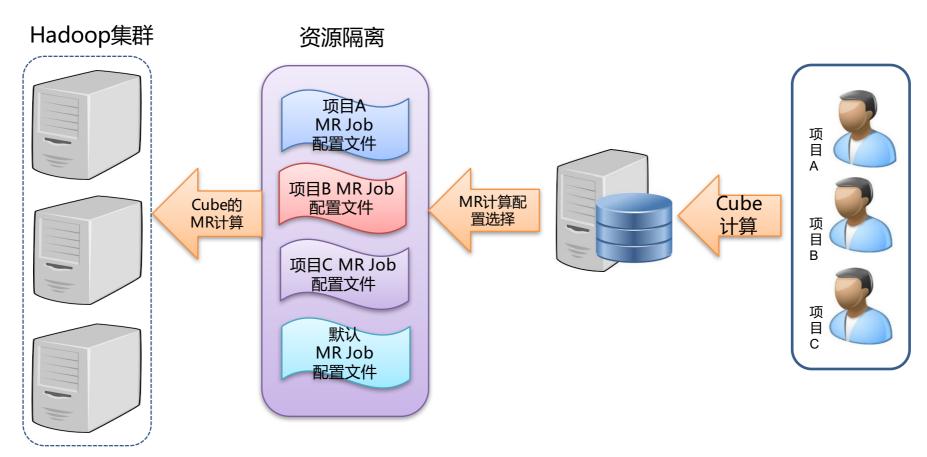
合并(merge)

每月中上旬自动合并上月所有 cube的分区



我们的深度开发 – 资源隔离

原生态暂不支持资源隔离,修改Apache Kylin代码支持项目级资源隔离





我们的深度开发 – 监控优化

数据拉取

依赖监控

Cube计算

任务监控

Hadoop、 Hive、HBase 集群监控

Apache Kylin 集群监控











Hadoop: 要Map和Reduce的 内存设置优化

HBase: Full GC优化, 关闭 MajorCompaction及压缩阈值调整,增大PermSize及ZK通信超时时间,等,有效减少 HBaseRegionServer随机Down机问题

Zookeeper: 增大ZK session

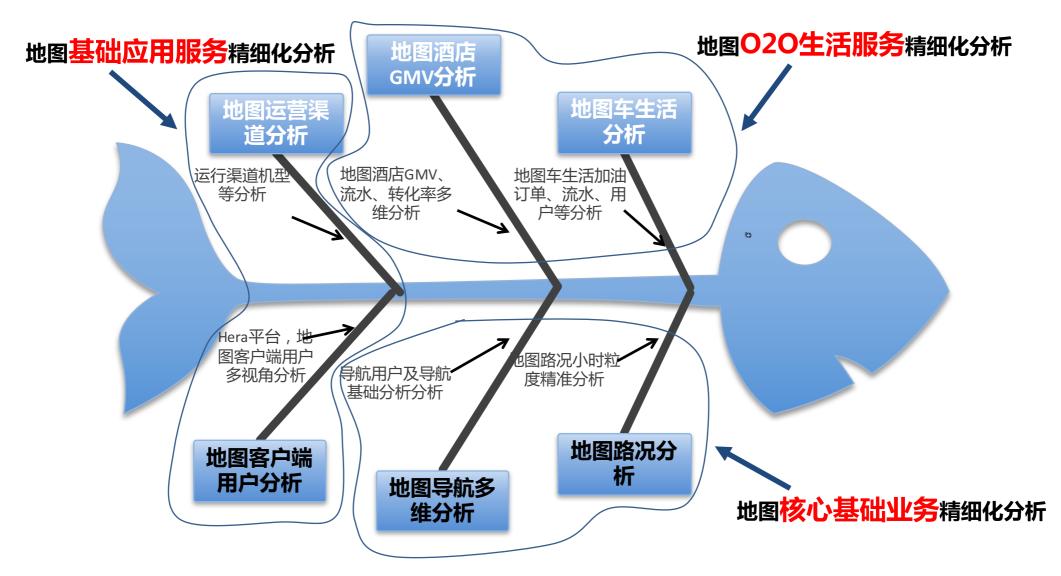
超时时间等



我们的项目实践



我们的项目实践 - 总体概览



共计50+亿行源数据,最大单表20+亿行记录



我们的项目实践 - 通用技术方案产出

1. Cube的查询性能主要因素之一 2. Cube数据以segment为存储单位, 即为Cube维度部分模型的设计, 主要设计计算、更新、分区合并等操作 **CUBE** 可根据同一维度的Hierarchy层次 ,目前已开发基于时间线端的自动化管 任务管理 关系及维度的元素的秩进行设计 设计 理逻辑。 不可累加 指标 3.对于某一维度下各元素值不 4. 留存涉及对历史多天数据的级联更 留存 互斥的情况下,向高纬度汇 新,影响数据规模大,性能降低,利 总累加出现数据膨胀错误, 级联更新 用数据旋转存储,前端辅助解决,仅 利用agg+detail cube结合 更新1天数据即可 路由算法实现 动态 5. 数据仓库端在biglog平台 做JOIN, 多维分析系统前端利 实际项目过程中归纳产出通用技术解决方案 **JOIN** 用Map实现JOIN 解决



Q&A 谢谢

